DOCUMENT RESUME

ED 042 799 TM 000 057

AUTHOR	Gorth, William P.; Wightman, Lawrence E.
TITLE	CAM Described for State Level Evaluation of Urban Education Projects.
INSTITUTION	Massachusetts Univ., Amherst. School of Education.
SPONS AGENCY	Charles F. Kettering Foundation, Dayton, Ohio.; New York State Education Dept., Albany. Div. of Evaluation.
REPORT NO	TM-20
PUB DATE	Apr 69
NOTE	62p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.20
DESCRIPTORS	*Academic Achievement, Achievement Tests, Behavioral Objectives, Comparative Testing, Computer Programs, Course Evaluation, Course Objectives, Curriculum Evaluation, Educational Objectives, Evaluation Needs, *Evaluation Techniques, *Measurement Techniques, Models, Post Testing, Pretesting, *Program Evaluation, Research Projects, State Departments of Education, State Programs, Test Construction, *Urban Education
IDENTIFIERS	CAM, *Comprehsnsive Achievement Monitoring

ABSTRACT

The Comprehensive Achievement Monitoring (CAM) model for project evaluation is compared with the usual classroom testing and pretest-posttest approaches. All three techniques are described and then examined in detail with respect to their advantages and limitations. The resources of project CAM in the areas of urban education and computer software are outlined. Three alternative methods for implementation of CAM by a state education agency and the practical considerations in implementation are examined. A bibliography on CAM is appended. (AE)

Project **C** omprehensive

A chievement

M onitoring

Technical Memorandum No. TM-20 April 1969

CAM DESCRIBED FOR STATE LEVEL EVALUATION OF URBAN EDUCATION PROJECTS

W. P. Gorth, Stanford University

L. E. Wightman, University of Massachusetts

William Gorth has been director of the CAM project since its inception two years ago, and has been responsible for the development of the techniques, evaluation designs and computer software to implement the procedures involved. This paper was prepared by Lawrence Wightman from material supplied by Mr. Gorth.

The writing of this report was sponsored in part by the Division of Evaluation, Alan G. Robertson, Director, New York State Education Department, Albany, New York.

The research and development of Project CAM is performed pursuant to a grant from the Charles F. Kettering Foundation to the principal investigator, Dwight W. Allen, Dean, School of Education, The University of Massachusetts, Amherst, Massachusetts.

Additional information may be obtained from:

William P. Gorth, Director
Project CAM
School of Education
Stanford University
Stanford, California [illegible]

or

Lawrence Wightman
Project CAM
School of Education
University of Massachusetts
Amherst, Massachusetts 010[illegible]

TM 000 057

Table of Contents

(Table of Contents, Continued)

Abstract

Effective project evaluation must rest on data which are comparable from one group or project to another and comparable across time, in order to measure changes in student performance accurately. Achievement related to individual goals of a project or set of projects must be clearly observable.

Usual classroom tests are not comparable across classes or over time, since each test is given only once, and only to one group. Usual project evaluation in urban schools is rendered almost invalid in some cases by high turnover in student population: the test is effectively nothing more than an immediate post-instruction test on the last few objectives taught. In addition, test items are not usually tied closely to stated objectives, especially when standardized tests are used. The CAM model of evaluation overcomes these problems by giving comparable tests on all objectives throughout the course. Computer-based analysis and reporting make possible the handling of data for large projects. Several plans are possible for state-level evaluation of urban education programs.

I. Introduction

The purpose of this report is to present an evaluation technique called Comprehensive Achievement Monitoring (CAM), and compare it with two other common types of evaluation. These are usual classroom testing, and pretest-posttest project evaluation. These three evaluation models will be related to the Guidelines for New York State Urban Education Program, and also, to some suggested additions to the Guidelines.

The CAM model is described in detail, and information and suggestions are given to aid in its implementation in evaluating urban education projects.

Present guidelines for evaluation. The Guidelines for New York State Urban Education Program list some general goals for the evaluation of Urban Education programs, but do not specify these goals in sufficient detail for individual project and program directors to collect data useful at the state level. The Guidelines suggest that project evaluation designs should require:

1) "data useful for decision-making at three levels--project, district, and state."

2) "systematic and objective accumulation of information (on) strengths and weaknesses" of each project.

3) data collection "prior to (the project's) inception, during its operation, and at its termination."

4) an "index (or indices) of each project's effectiveness."

5) information for making "decisions (on) continuing or modifying initial projects."

It is difficult to obtain an index of effectiveness, to observe an increase or change in academic achievement, if the measures of achievement at one time are not directly comparable with the measures of achievement at another time (the notion of parallel criteria or parallel test measures). It is therefore desirable to plan at the inception of a project a systematic pattern of frequent data collection, using similar and comparable instruments. The Guidelines call for information to be gathered throughout a project, although the specific types of data, and the actual schedule or design for the data collection, are not enumerated in detail. However, the frequently-used posttest or pretest-posttest designs for project evaluation do not fulfill the implications of this goal.

Suggested modifications for the Guidelines. In order to increase the usefulness of information for project and state level decision-making, evaluation designs should require:

1) common instruments to evaluate similar projects.

2) similar schedules of data collection.

3) results reported in sufficient detail to make judgments about pacing and sequencing of instructional procedures.

4) feedback of results during the project to improve instruction.

5) a data bank of evaluation results upon which to base long-range policy decisions.

6) financial data to combine with evaluation data to begin a cost-effectiveness analysis.

Specific recommendations must be made at the state level to insure that projects systematically collect information which will fulfill the

requirements of evaluation. The information should include:

1) estimates of performance related to the specific objectives of the project.

2) systematic and repeated pre-instruction testing of performance on each objective.

3) immediate post-instruction testing of performance on each objective.

4) retention testing of performance on each objective.

Care must be taken that the data gathered will allow a continuous, longitudinal appraisal, and that it will be collected in such a way as to keep to a minimum the effects of attrition.

II. Components of CAM

CAM is a procedure for testing achievement on every objective of a course, at frequent test administrations throughout the course. At each test administration, performance on objectives not yet taught is pretested, performance on objectives just taught is immediately posttested, and performance on objectives taught earlier in the course is measured for retention. Parallel test forms, comparable in difficulty and content, are all used at each test administration, but each student receives a particular form only once during the course. Each form typically has an item for each objective. Each item is used on only one test form. The function of a particular item changes in relation to the time at which its objective is taught. Testing may take place at regular intervals (e.g., every two weeks) or at the end of certain instructional units. Computer based analyses and reports are available within a few days of data collection.

Specification of objectives. The most fundamental preparatory step for the use of CAM is the specification of the objectives to be evaluated, in testable, behavioral terms. Objectives may be categorized according to numerous dimensions, and possibly organized into instructional units. Written objectives for a variety of closely related projects or courses may be collated and pooled. It is then possible to identify and select for evaluation those objectives which are common to several projects, and those that are unique to a project. Objectives are typically related to achievement; however, CAM is equally suited to measuring changes in attitudes or

perceptions. The pool of objectives is called an objective bank, and a computer program is available to handle the large amount of data involved.

Test items. The second step toward the use of CAM is the construction of test items. Every item is tied specifically to a single objective, and multiple items are constructed for each objective. All items, keyed by objectives, may be stored in a computerized item bank, ready for sampling or available for revision.

Construction of test forms. The number of test forms, or monitors, must at least equal the number of test administrations planned. Tests are made parallel in content by using the technique of stratified random sampling. Forms are also randomly comparable in difficulty. If an item analysis can be run (perhaps on a pretest or an earlier version of the course) for indices of difficulty and discrimination, the forms may be made more exactly comparable in difficulty.

Monitors are intended to be short tests, perhaps ten to thirty items. Whether or not a single form covers all objectives for a course is a function of the proportion of objectives to items-per-form. It may be necessary to randomly sample (without replacement) the objectives, before doing the same on the test items for each selected objective. This technique of sampling must insure that, across forms, all objectives are equally represented. The same consideration holds when items-per-form exceed the number of objectives; in this case, some objectives may be represented by more than one item on some forms.

Student test groups. Students are divided into test groups in order to use all test forms at each administration. Test groups are best constructed using random sampling of strata of students based on ability or

prior achievement in the subject. This assures that each group has a range of students which gives representativeness to the data for each test form.

It is most desirable, for several reasons, to include every student in every test administration, and when set up this way, CAM has been found to be a satisfactory substitute for usual classroom testing. However, it is possible to use only a sample of the student population, especially if the number involved in a project approaches one thousand or more. Many different sampling designs are possible; some are shown in Figure 20.1. Using the total student population in one test group is the design for the conventional project evaluation. Unequal-sized test groups may sometimes be an administrative necessity.

Test administrations. Test administrations may coincide with the completion of instructional units, or they may be set at regular intervals throughout the course. The latter has advantages in terms of ease of administration, and comparability of results from similar courses taught at different schools.

Appended package tests. It is possible to add a section to any monitor, and have the results incorporated with the rest of the CAM data. This feature lends flexibility in that, should a specific diagnostic test seem desirable at any point, the data can easily be assimilated.

Data analysis and reporting. Output from the computer programs is as follows:

For individual students

After each administration:

1) total score on that and all previous administrations.

2) a graphic presentation of the above.

3) a right-wrong indication for each item on the monitor, coded by the objective represented.

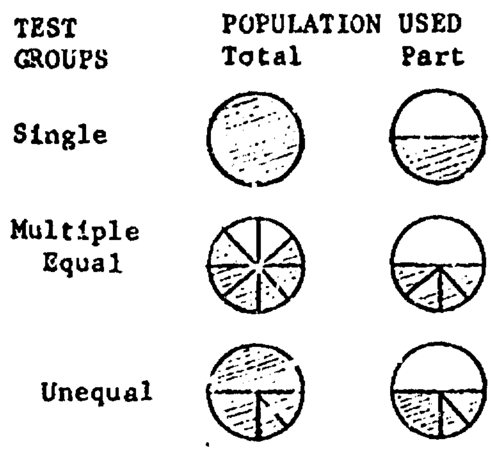


Fig. TM-20.1. Some patterns for sampling students. Circle represents student population. Shaded part represents sample tested at one administration, within which divisions indicate test groups.

At the end of the course:

4) average scores, across all monitors taken, on items categorized by use into three groups--pretest, immediate post-instruction and retention of varying lengths of time.

For whole group or subgroups (e.g., one classroom; highest and lowest quartiles)

After each administration:

1) percent answered correctly out of all items across all monitors, for each objective.

Periodically, as desired (e.g., every 3-5 administrations):

2) trend data, or achievement profiles, for total score and for each objective.

At the end of the course:

3) same as number 4 under individual students.

4) item analysis (using whole group only), treating each item in three separate ways, by its three functions--pretest, immediate post-instruction, and retention measure.

Data are analyzed, and reports printed, by computer; results are usually available within a few days of each test administration. Data can be collapsed in various ways, to be most useful to students, teachers, project directors, or state evaluators.

III. Usual Classroom Testing

Description

The usual classroom testing situation concludes the following sequence of events: first, a set of objectives is specified for a limited instructional period, usually from one to four weeks; second, an instructional treatment is devised and administered to the students; and lastly, a test at the close of the instructional period is administered to measure the extent to which the objectives taught during that period have been achieved.

Students' achievement on material taught during instructional period one is tested at test administration one. Achievement for period three is tested at administration three, and so on, throughout the course.

There is usually a "final test" administered at the end of the course, for which there may be varying amounts of review offered. Sometimes major tests are administered at other times during the course, e.g. just before report cards are issued.

Strengths

Flexible weighting. There is great flexibility in the relative emphasis accorded various objectives during the year. Decisions may be made at any time; content may be added, dropped or modified. The testing is tailored to the content as the course progresses.

Individual student testing. Usual classroom testing can yield diagnostic data on individual student achievement, on the few specific objectives which have just been taught.

Tests related to objectives. Usual classroom testing may meet the criterion of close relationship between objectives and test items, when the school program is defined in behavioral objectives, and the teacher makes some effort to relate the items directly to the objectives.

Limitations

No pretesting. There is usually no pretest information on students' prior achievement on any objective. Teachers usually assume that student achievement is due solely to the instruction given them in class. Furthermore, they do not know whether learning one objective has affected understanding of another objective. Students may also have experiences in other courses, or outside of school, either before or during a course, which contribute to their understanding of various objectives, whether or not they have been taught yet.

No test of retention. There is no information on students' retention of objectives which have been taught earlier in the school year, except in the event of some sort of major test. At that test administration, the interval between time of instruction, and time of test-of-retention, is different for every objective taught. The interval may span almost a full school year, or be only a week or two. There is seldom any data attached to such test results about the date of instruction on a given objective.

No comparison of student achievement over time. It is very difficult to compare students' achievement from one point in time to another, because at each test administration, an entirely different test is used; there is seldom any overlap in content, and the overall dfficulty can vary enormously from one test to another. The only possible comparison of achievement

from one time to another must use a student's rank order in his class. This still leaves no way to examine changes in a total class's achievement over time.

IV. Project Evaluation

Description

A frequently used strategy for evaluating projects is to administer an extensive achievement test at the conclusion of the project. This may consist of a test, or battery of tests, sometimes composed specifically for the project, but usually prepared and distributed commercially, e.g. standardized achievement tests.

There is sometimes a pretest administered before the start of the project, which is either the same as the posttest, or an alternate form of it, but presumes to measure the same objectives.

Limitations

Deficient immediate post-instructional testing. In terms of immediate post-instruction achievement, the usual project evaluation measures only the objectives taught at the very end of the project in a way similar to usual classroom testing (i.e., immediately following the instructional treatment). This means that project directors do not have information on the direct effect of instruction immediately after students have been exposed to it.

Tests of retention. The interval between the teaching of an objective, and the end-of-course test, varies for each objective. Such intervals range from a week or two, to a full school year. Therefore, an estimate of achievement based only on a posttest is an aggregate of immediate post-instruction achievement, short-term retention, and long-term retention. This composite score may be made up of several subscores,

but such subscores still do not indicate much about the time interval since instruction.

No comparison of scores. There is no need to discuss comparability of scores from one time to another if the testing is done at only one point in time. Pretest-posttest problems are discussed below under sample attrition.

Test items not specific to objective. In posttests which are designed to cover an entire course at only one administration, there is great variation in the specificity with which test items have been matched to the objectives of the course. This problem is especially apparent when standardized achievement tests are used, where general subscores are roughly matched with the stated objectives of the project. When only standardized tests and materials are used in a post-project evaluation, there is a definite lack of systematic information about the achievement on specific objectives in the program.

Inappropriate weighting. In giving one large posttest, especially a standardized test, the problem of weighting of objectives presents itself. A variety of objectives could be poorly measured while other objectives are heavily emphasized. It is likely that the intended pattern of emphasis in the course will not be reflected in the evaluation instrument.

Test not comprehensive. Not only will there be too little emphasis on certain objectives, but it is possible that some objectives will not be measured at all. Lack of comprehensiveness in an evaluation technique is a serious shortcoming.

Problems of sample attrition. All of the above weaknesses in the usual project evaluation design are relatively unimportant when compared with the most serious problem of all: the turnover of students, which in urban

schools is extremely high; i.e., it sometimes equals 100% of the enrollment. Those students who were pretested before the program, and received the early segments of instruction, are simply not there at the time of the posttest. Effectively, this reduces the hard data to a posttest on students still enrolled in the project during the final week, even if a pretest were administered. Therefore, the results may represent very little more than immediate post-instruction testing on the objectives taught just before the posttest. Pretest information, if it has been gathered, relates only to the incoming abilities of a sample of students roughly similar to that available for posttesting. The assumption is made that students coming into the project are similar to those leaving it, but the data cannot be used statistically in analyzing changes in student achievement, since change should only be measured for individual students or identical groups of students. A pretest-posttest design for evaluating projects does not seem to be reasonable for the types of programs usually encountered in the Urban Education Program.

Advantages

In the light of the serious limitations of posttest and pretest-posttest evaluation designs, it may be ironic to point out one seeming advantage, but in the interest of perspective, a single posttest (and perhaps pretest-posttest) does cost less than a more effective and complete evaluation system such as CAM. There is a minimum of clerical and administrative work needed in actually giving the test, and if a commercially available test is used, it may simply be purchased; no staff or time is needed to develop a test tailored to the objectives of the project. What little analysis on results can be done, is relatively easily accomplished.

V. Capacity of CAM

Advantages

This section highlights the comprehensiveness, the precision, and the timeliness of information available from the CAM model, as contrasted with the usual classroom testing and project evaluations. It should be emphasized that the validity of the estimates of group achievement available from CAM is comparable with, and in many respects superior to, that of the more familiar techniques.

Specificity of objectives. Any project, no matter how it is to be evaluated, can call for a high degree of specificity of objectives; CAM, however, rigorously prescribes and requires such specificity. It is the base upon which the detailed testing, analysis and feedback of the program rest.

Specificity of objectives allows similar projects to pool and match their objectives. What is common to all projects, or to several, is readily observable, and provides a meaningful, detailed comparison. Objectives unique to individual projects can pinpoint actual differences concretely and precisely.

Test items tied to objectives. Each test item is constructed to measure achievement on a particular objective. Therefore, test data always relate to definite objectives, rather than aggregates of objectives; this allows evaluation procedures to be matched with specific goals of the project. In this respect CAM differs significantly from conventional project evaluations, where standardized materials are used, which have not been closely tied to the specific objectives for a project.

Modification of projects. Conventional project evaluation may provide some criteria upon which to base one kind of decision about an existing project: "drop it" or "continue it." These criteria are global rather than related to specific contributions of the project. Perhaps one of the most valuable characteristics of the comprehensive achievement monitoring model is that it is able to provide information upon which to make specific recommendations for retaining strong components of a project, and modifying weak ones. No project is as effective as possible, as set up at its inception; therefore, a far more pertinent decision about it, now possible with the CAM model, is "drop" or "continue with these modifications."

Data more valid. If there is time on a test for one question for an objective, then estimates of group achievement on that objective will be more valid if a variety of questions is used across the group, rather than the single question typical of both classroom tests and project evaluation. It is important to note that the increased validity and comprehensiveness calls for no sacrifice in the economy of data collection, since each student need still answer only one question.

Pretest of all objectives. All objectives are pretested before any instruction has been given. First, it is important to know whether students have already acquired information or skills from outside sources, so that the project need not lose students' interest by covering material that they can handle already. Secondly, an index of effectiveness must ultimately be an index related to change in student achievement, attitude or perception. In order to document change, it is necessary to have at least two comparable measurements of the same characteristic, taken at two different times.

There is reason to continue pretesting on objectives to be taught later

in the project, because outside learning experiences, or interaction between material taught early in the project and that scheduled to be taught later, may both very reasonably cause changes in performance during the project. This may lead to alterations, either in the sequence of instruction, or the amount of time spent on certain objectives. When the level of achievement rises on an objective not yet taught, it may be closely related to material just taught, in which case, instruction in the later-scheduled unit could be moved up to take pedagogical advantage of the relationship. Another possibility is that, without changing the sequence, certain instructional units might be condensed, and the pace of instruction stepped up. A single pre-course test, will not provide information for making the above decisions.

Immediate post-instruction test. The usual classroom test covers only material just taught. CAM estimates of group achievement on just-taught objectives are comparable to those available from classroom testing. The numbers of students usually involved in projects makes it possible to test each objective with a substantial variety of items, without lengthening any one form of the test.

Continual measure of retention. Since objectives continue to be tested after they have been taught, throughout the rest of the course, there is a continual test of retention. Intervals between "teach" and "test" times are of varying length, and can be matched for precise analysis. For example, it would be possible to measure retention spanning approximately six weeks on all material of a course except what is presented during the last month or so. Therefore, estimates of achievement can be systematically made for each of the instructional units after a specified interval.

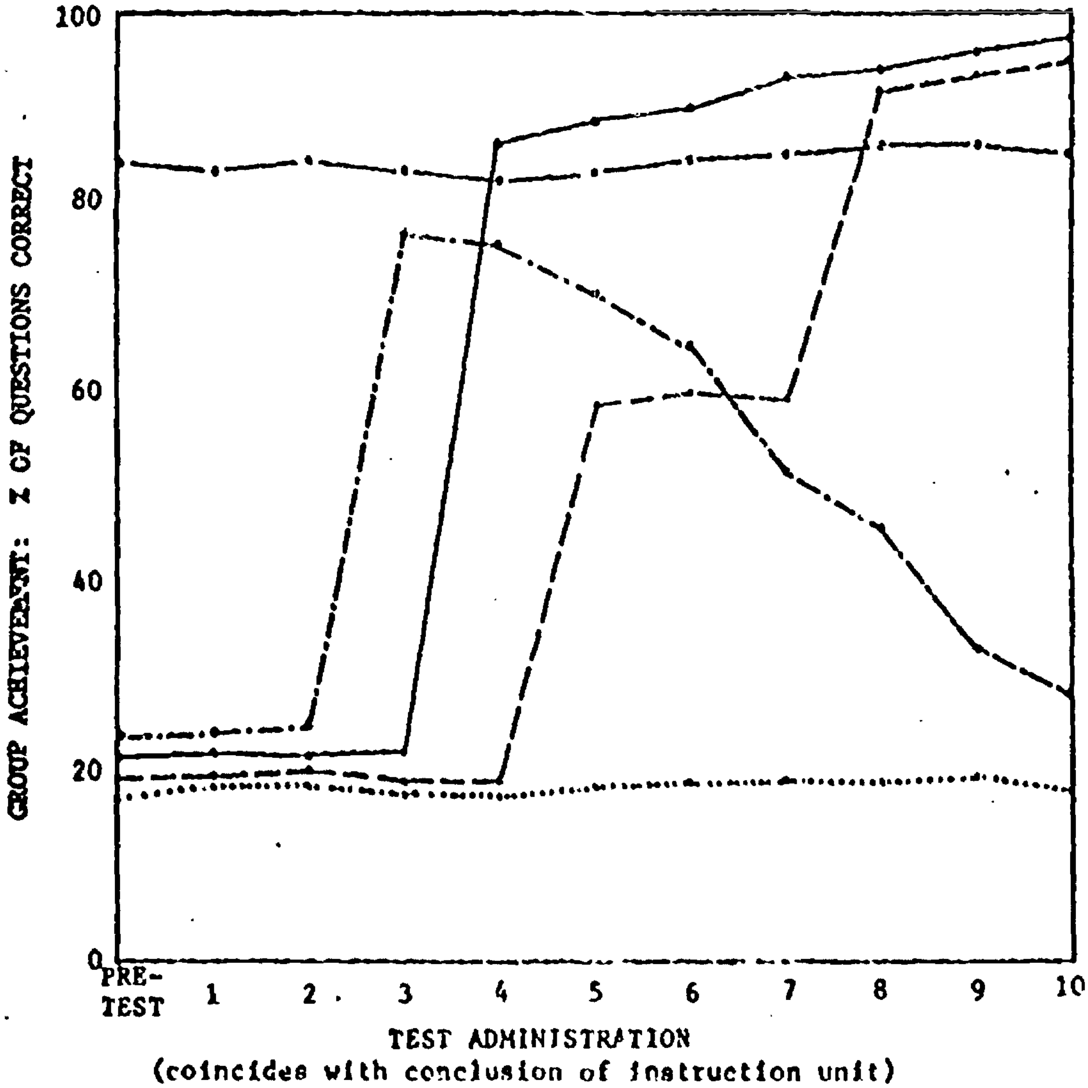
One of the major indices of an effective program in urban education would be the retention of material learned during the project. This must be systematically and continuously measured, and the variable of time-of-retention must be included in any analyses of data. The usual classroom testing model does not permit this type of analysis because it does not provide this type of data.

Tailored review. Information on retention can have another side benefit: a possible saving of teaching time. Teachers often set aside class time before major tests for review. However, if it can be shown that students have not forgotten certain objectives, there is no reason for repeating the instruction.

There are as yet no data to show whether the use of monitors, containing questions on previous materials, can actually aid in student retention.

Achievement profile. There are comparable data on achievement for every test administration. This makes it possible to plot students' achievement on any given objective (or group of objectives) for the entire course. This plot, called an achievement profile, gives a graphic presentation of the changes in group achievement throughout the course. This achievement profile is a unique characteristic of the information available from the CAM model and is very useful in describing and reporting results of course and project research.

Figure TM-20.2 presents hypothetical achievement profiles for five objectives from a course. Brief comments below the graph give possible interpretations. It is obvious that achievement profiles provide a wealth of information, at whatever point in the course they are drawn. On the pretest in the foregoing example, all objectives except number 2 show achievement at the chance level, or about 20% (five-option multiple-choice items). Several decisions could have been made after test administration one:



Obj. 1: taught, but students did not learn; with rapid feedback, could be corrected with change in instruction.

Obj. 2: previously known and not taught; without pretest, this looks like student learning.

Obj. 3: taught and learned, but forgotten

Obj. 4: well taught

Obj. 8: appears related to objective 3, because achievement increases when 3 is taught.

Figure TM-20.2. Hypothetical achievement profiles of group achievement on five objectives.

1) Objective 1 was not learned--reteach it in some other way; 2) Objective 2 has tested high on both the pretest and test administration 1--it would be safe to skip instruction in this objective. After test administration 5, two other decisions might have been made: 1) Achievement on Objective 3 seems to be slipping--review is needed, preferably soon; 2) Objective 8 seems closely related to Objective 5--perhaps it should be taught now instead of later.

Other variations of achievement profiles are available. Figure TM-20.3 shows total test scores averaged by quartiles of the class. Other subgroupings of students are possible. Any desired subgroupings of students is possible. Figure TM-20.4 shows the profiles of four students (total scores) across the semester. The computer programs are general so that profiles for any combination of items for any subgroup of students may be obtained.

Relative immunity to attrition. To gauge the effectiveness of a project, change in student achievement must be documented; measurement must be available for an objective both before and after it is taught, and the measurement must be on the same students. The usual project evaluation measures achievement, at best, only at the beginning and end of a project. The rate of turnover of student population in urban schools is notoriously high. However, the unique capability of the CAM model allows comparable measures of achievement at multiple points in time. The shorter the interval between tests, the greater the likelihood that students enrolled in the project at test administration A, are still enrolled in the project at test administration B. Much data can be salvaged through this technique, but it is irrevocably lost when the usual pattern of project evaluation is followed. This is illustrated in Table TM-20.1.

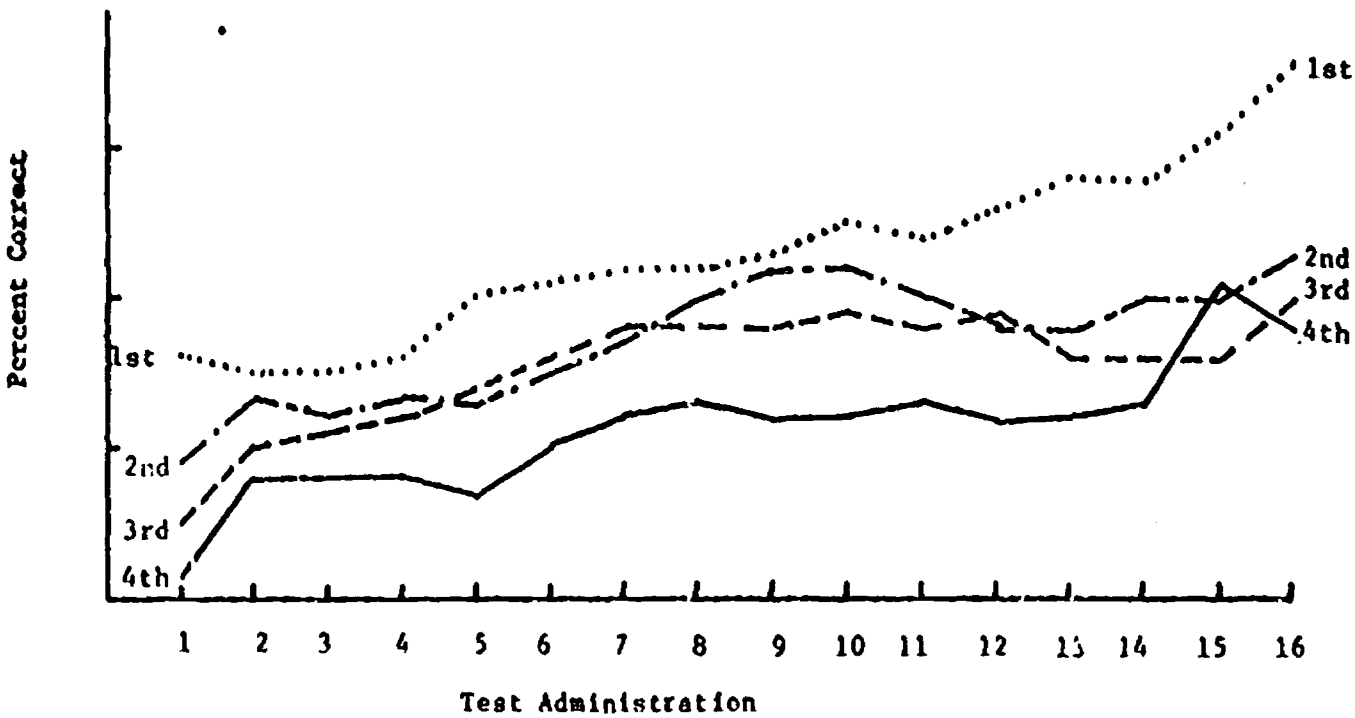


Figure IM-20.3. Achievement profiles of quartiles of a class by test administration. (Percent correct for each period is the mean percent correct of that period and the two adjacent periods.) Students were divided by their scores at Administration Intervals were about two weeks long. Administrations 1 and 16 were pretest and posttest, which were taken by every student. The other administrations involved a 30% sample of students.

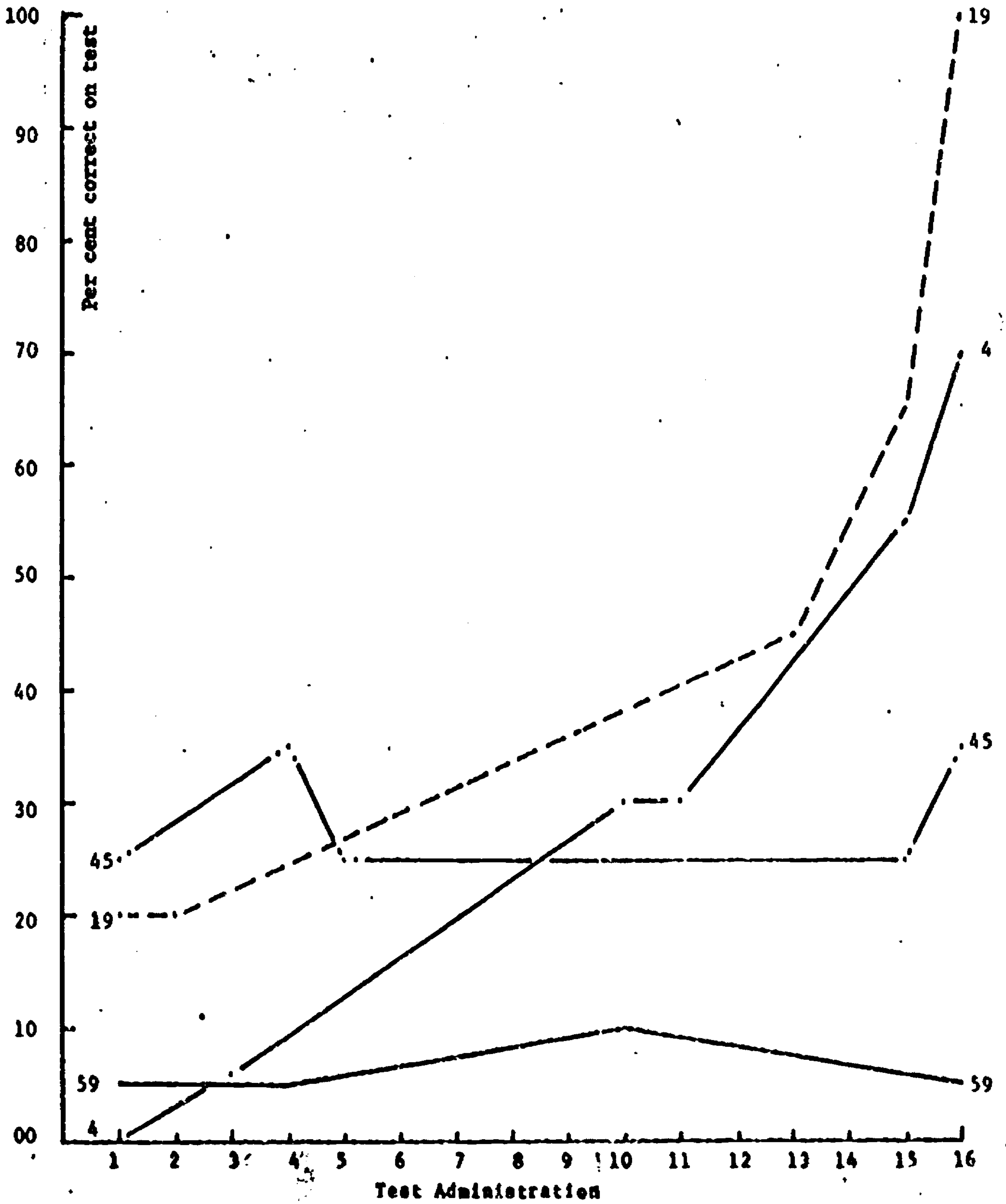


Figure TM-20.4. Achievement profiles of four students; total test score by test administration. Intervals were about two weeks long. Administrations 1 and 16 were pretest and posttest, which were taken by every student. The other administrations involved a 30% sample of students.

TABLE TM-20.1
Evaluation Cost Analysis:
Attrition of Hypotehtical Student Sample
as Related to Percent of Evaluation Data Retained

Group Entering After *	Enrollment				
	beg. September	mid November	end January	mid April	end June
September	1000	800	700	600	400
November		200	175	150	100
January			125	110	70
April				140	90
					340

Evaluation Design	Percent of Data Usable	
	For Period	Percent
Pretest-Posttest	September-June	40
CAM	September-November	80
	November-January	88
	January-April	86
	April-June	66
	CAM AVERAGE	80

*First figure in each row represents entering students; subsequent figures in that row represent students in that group still in school, for whom are useful.

Continuous data available. Data are available from every test administration. It is possible to look at group achievement on a single objective, groups of objectives, or total content of a course, though this last is generally less useful. Data can be summarized in a variety of ways, through the use of selected computer programs now available. Desired data are always available within a few days for decision-making; it is not necessary to wait weeks or months for meaningful analyses. Many evaluation systems are not able to analyze and report results with sufficient speed and organization to make the information most useful to its recipients. Analyses can be tailor-made for project directors or state evaluators.

One economic advantage of periodic feedback is that a project need not continue to its end to discover, after all funds are spent, that the goals of the project have not been accomplished. Modifications can be made in the program if student performance does not move in the expected direction.

Indices of effectiveness. One of the most pressing problems in the relative evaluation of projects is comparability of data from one project to another. Such a measure might be called an index of effectiveness.

For decision-making at the state level, where several similar projects may be competing for continued support, courses must be compared on whether students achieve the behavioral objectives which they have in common. Since common objectives, and matching test items, can be selected from a large pool, data from several projects can be compared with precision. It is important to note that items (or objectives) that are different from one project to the next may have different levels of difficulty, and therefore comparison of projects using scores on these items may be distorted.

Project CAM has a program that calculates for individual students, by

objective, scores for items which were pretest items, items which were immediate posttest, and for items which were retention measures (see Table TM-20.8; for more technical data see TM-6, TM-11, TM-12). If monitors were either wholly or largely the same for several projects, it would be relatively easy to compare projects on these measures. Additional measures of effectiveness might include analyses of group learning curves, or more formal trend analysis or time series regression. These measures could be applied to achievement data, attitudinal data, attendance data and/or social (process) data.

One further step in analyzing effectiveness of projects, once a measure of effectiveness is established, would be to analyze effectiveness in relation to financial information on each project. Rudimentary cost-benefit analyses would show which projects having comparable objectives were most effective <u>per dollar</u>. This would seem to be the ultimate purpose of project evaluation.

<u>Important Considerations in Use of CAM</u>

<u>Good items needed</u>. The development of an item pool for CAM takes some special skills and considerable staff time. As in all large evaluation projects, items must be of high quality; with CAM, there must also be <u>more</u> items. Ideally, an item analysis should be run, to insure that all monitor forms are of equal difficulty. In the absence of such an analysis (and there often is not time or opportunity for it before starting a program), it is doubly important that the staff assigned to item construction be skilled and experienced. Forms will be randomly equivalent in difficulty.

<u>Staff needs</u>. In addition to the initial construction of test items, other components of CAM also need the attention of highly skilled staff

members. Items must be selected for each specific project; monitor forms must be carefully devised with proper sampling procedures; there may be student sampling procedures to accomplish; interpretation of test data must be done correctly; analyses and summaries call for certain technical skill.

In addition to academic and technical staff needs, there is some clerical work involved in keeping track of monitor forms, students, administrations, and results.

Sampling problems. When it has been decided, for reasons of economy, to sample the student population rather than test everyone at every administration, certain problems can arise.

Students are accustomed to feedback from most tests they take; they may become indifferent to a series of tests which seem sporadic and yield no reinforcement. Also, a student may be resistant to taking a test if he knows that some of his friends are not.

When CAM is set up to test comprehensively (every student on every objective at every administration) it has been found to be an adequate substitute for classroom testing. Obviously, if students are being sampled, classroom tests for diagnosing individual student achievement must be given in tandem with the monitors.

Attrition problems. There is no way to solve completely the problems of student sample attrition. Achievement on specific objectives will show up, but if there is a rapid turnover in the student population, total scores *may* rise very little. Some of the information about transfer of learning from one objective to another may also be distorted.

Low total scores. Total scores for a project must be interpreted carefully. If scores are low, several factors may be at work. Attrition problems

can account for "learning and forgetting" achievement profiles. Another problem may be the overall difficulty of items, which of course are not modified during the year. Careful analysis can obviate these difficulties; it is important that incorrect interpretation not be the cause of unfavorable comparison to other projects.

Summary Comparison of Three Evaluation Models

The amount and quality of information available from the three models of evaluation described above will serve to summarize the characteristics of each.

Comparison of matrices (amount of information). CAM yields more information than either the usual classroom testing or conventional project evaluation. The pattern of data resulting from each model may be fitted into a matrix, in which the rows indicate all the objectives or instructional units of the course, and the columns represent the possible test administrations during the entire project. A cell of the matrix which is filled in, represents an estimate of achievement for that objective or unit, at that test administration.

The usual classroom testing pattern is illustrated in Table TM-20.2. The diagonal line of X's represents the series of separate tests on each instructional unit, each given at separate administrations. The column of X's at the last administration indicates a final test, presumably covering all the units of the course.

Table TM-20.3 illustrates graphically the lack of information available from the usual pretest-posttest project evaluation. This illustration makes the assumption, not necessarily well-founded, that a single test does in fact provide information about every instructional unit.

TABLE TM-20.2

Usual Classroom Testing: Estimates of Achievements Available for a Group of Students by Unit and Test Administration

	Time					
Unit	1	2	3	4	...	T
1	X					X
2		X				X
3			X			X
4				X		X
.					X	X
.						
U						X

TABLE TM-20.3

Pretest-Posttest Project Evaluation: Estimates of Achievement Available for a Group of Students by Unit and Test Administration.

	Time					
Unit	1	2	3	4	...	T
1	X					X
2	X					X
3	X					X
4	X					X
.	X					X
.						
.						
U	X					X

It is readily apparent in Table TM-20.4 that CAM makes available data on group achievement for all of the objectives specified for a course, at each time of testing. This comprehensiveness of the data provides the necessary information for the variety of purposes discussed earlier in this section. It is easy to see how CAM contrasts with the other models of testing, where information is generally available either on a few of the objectives, or as a composite score for all objectives, at a single time.

Quality comparison. Table TM-20.5 displays seven types of information, and estimates their quality as provided by each of the three models.

Conventional project evaluation is fair to poor on all of the dimensions described. These shortcomings are inherent in the use of single tests at one, or perhaps two, points in time. A single test long enough to provide detailed information about student performance on a large number of objectives is fatiguing and therefore less valid than short tests. One long test excludes systematic pretest, immediate post-instruction, and detailed retention information. Attrition takes a heavy toll of a pretest sample. Feedback is limited to a post-mortem on the project's strengths and weaknesses.

Usual classroom testing provides for the measurement of performance on specific objectives on an immediate post-instruction basis. By repeated testing, the effects of attrition may be minimized. If usual classroom testing data were collected across similar projects after similar objectives had been taught, extensive information would be available for comparing projects. However, an accurate comparison of projects must also include pretest and retention information. The former is used to adjust for incoming aptitude and achievement differences in students, and the latter for long-term retention, or payoff of the project. Neither of these is specifically available from

TABLE TM-20.4

Comprehensive Achievement Monitoring Evaluation:
Estimates of Achievement Available for a Group
of Students by Unit and Test Administration

	Time					
Unit	1	2	3	4	...	T
1	C	C	C	C	C	C
2	C	C	C	C	C	C
3	C	C	C	C	C	C
4	C	C	C	C	C	C
⋮	C	C	C	C	C	C
U	C	C	C	C	C	C

TABLE TM-20.5

Quality of Information
Available from Three Evaluation Models

Information	Model: Usual classroom testing	Usual project evaluation	Comprehensive Achievement Monitoring
Pretest of objectives	*	*	***
Evaluation specific to objectives	***	**	***
Immediate post-instruction testing	***	**	***
Evaluation of retention of objectives	*	**	***
Comparability across time	*	*	***
Achievement profiles			****
Continuous feedback	**	*	****
Immunity to sample attrition	***	*	****

NOTE: Quality of information rated as excellent (****), good (***), fair (**), poor (*), and not available (blank).

classroom testing. Feedback occurs frequently during the project, but provides information about only one instructional unit at a time.

Comprehensive Achievement Monitoring provides information for project or state evaluation comparable, or superior, to the other evaluation models. Its superiority lies in the areas of particular importance to project evaluation: systematic pretests and measures of retention of objectives. Attrition which can easily invalidate the results of an evaluation of urban education programs, is accommodated in an automatic way. Feedback can be provided continuously and comprehensively (as called for in the Guidelines) so that the projects can be critiqued and adjustments made before their end.

VI. Resources of Project CAM

Staff

Project CAM has been involved in a variety of courses and grade levels: in the process, a staff has been assembled with a valuable range of skills in using CAM procedures.

Experience in urban education. Several CAM staff members have initiated, or worked with, educational projects for the disadvantaged. They have worked with Headstart programs, occupational guidance in a ghetto, a "drop-out prevention" project, and individualized instruction in inner city high schools funded with Title I monies. The combination of project activities in another context, and participation in CAM, has allowed these staff members to develop an expertise well-suited to urban education evaluation.

Perhaps the most difficult aspect of evaluating urban education projects lies in establishing rapport with project teachers and students. They must be approached by competent and sensitive evaluators; and the validity of the data collected can be seriously compromised by indifference or lack of understanding on the part of teachers or students. The ideal situation would be to have a few highly trained, skillful and sympathetic evaluators, who could go into schools involved in a project. They could explain the evaluation system, work out specific problems that might arise, and train one or several local staff members to administer the monitors correctly.

Experience in design and analysis. Every project has characteristics which are unique to it: objectives, pacing or sequencing of instruction, class organization of students, etc. Each of the courses using CAM has had a different structure: teachers singly or in teams; traditional, individually paced, or individually prescribed instruction. Using the basic CAM design, a specific evaluation design was developed to suit each course. Statistical analyses were performed to provide each teacher with the maximum amount of information concerning the special features of his course. Members of the CAM staff have demonstrated their flexibility in accommodating variations in course (or project) requirements.

Table TM-20.6 shows the range of courses for which CAM has been used, or is presently being planned. Table TM-20.7 gives some detail for courses where CAM is in current use.

Computer Software

In order to be most useful to project directors and state level decicion makers, the steps of an evaluation must be accurate, flexible and timely. CAM procedures are capable of such accuracy and speed, because of the development of related computer software, i.e., computer programs, to handle the large amount of data involved. A summary of available programs follows.

Objective bank. This program stores the text of objectives, and allows them to be revised. Objectives may be categorized in a variety of dimensions and may be retrieved through any category. The program will print lists of objectives.

Item bank. Items are coded by their objectives and/or other indicators; they can be stored (with correct answer keyed), edited, selected at random within objectives, and printed in sets to produce randomly parallel test

TABLE TM-20.6

Subject and Grade Levels for which CAM is Operational or Planned

Subject	Grade level					
	Pre-high school	9	10	11	12	Post- high
Science	P	0	0	P		0
Mathematics	P	0	0	0	0	0
English			0			
History	P		0	0	0	
Vocational		P	P	P	P	0

TABLE TM-20.7

Summary of Courses Monitored by CAM, 1968-69

Course	Relevant Data					
	Subject matter and grade	Method of instruction	Course size	Testing frequency	Number of forms	Items Per form
DH 110	English, 10th	Individualized Instruction (I.I.)	150	Every 5 wks	10	20
DH 210	History, 10th	I.I.	140	Every 4 wks	10	20
DH 321	Biology, 10th	I.I.	140	Every 4 wks	9	15
DH 410	Math, 10th	I.I.	110	Every 4 wks	10	20
HS 420	Math, 11th	Traditional	130	Every week	14	9
JN 402	Math, 9th	I.I.	280	Every 2 wks	15	14
KA 240	History, 11th	Traditional	130	Every 4 wks	12	24
KA 442	Math, 11th & 12th	Traditional	150	Every 2 wks	9	14
ML 205	Math, 9th	I.I.	150	Every 3 wks	10	30

forms. CAM procedures may call for 800 items to be distributed across 40 test forms, for one of numerous related projects; the computer program can accomodate many thousands of items, belonging to any number of objectives. More information on the item bank program is available in Technical Memorandum 14, and in Gorth and Grayson (1969).

Data bank. This is a "tape building" program; i.e., it continually adds information in the correct place on one tape, which is the data storage vehicle. The program inputs, edits, collates and systematically stores information on each student, and the instruction on each objective; in addition, for every monitor of the set, each item is noted with its correct answer, and categorized in up to fifteen dimensions (e.g., Bloom's taxonomy, 1956), one of which is the objective to which the item is tied. Data on each student include:

1) his name and available background information, personal and academic.

2) the sequence of tests he is assigned, which is independent of other students' sequences.

3) the date he completes instruction on each objective.

4) at each administration, the date, test form and specific responses he gave.

It is evident that no data are lost, and the format of this program allows flexibility: instruction may be individually paced and/or prescribed if desired, and students may be individually scheduled for testing. For more information, see Technical Memorandum 7.

Test data analysis and reporting. As soon as test data are put into the data bank, they are available for analysis, along with all the previously stored data. This program, therefore, calls on the data bank for its raw material. Reporting includes: 1) group achievement on each objective over time; 2) total test score for the current and all previous administrations, for individual students; 3) right-wrong information on each item, keyed by objective, taken by an individual student of that administration. Figures TM-20.5, TM-20.6, and TM-20.7 show sample outputs of this program.

Achievement profiles. This program provides a detailed analysis of performance, for 1) groups of students, e.g., group achievement on a single objective, or 2) groups of items, e.g., trend in total score for each student, or 3) combinations of 1) and 2). It is a graphic presentation of achievement across time. Sample output can be found in Figure TM-20.8, and more information is available in Technical Memorandum 12, and Gorth, Grayson and Stroud (1969).

Item analysis. Not every student receives every item after the corresponding objective has been taught, and there is no one test score which can provide a valid criterion against which to measure the difficulty of a given item; therefore, the usual procedures for obtaining difficulty and discrimination indices are not applicable to CAM data. A new procedure was developed (Technical Memorandum 6; Lindeman, Gorth, and Allen, in press) which allows three separate treatments of each item, by its function: pretest, immediate post-instruction or measure of retention.

Implementation by a computer program is discussed in Technical Memorandum 11. Sample output is abstracted in Tables TM-20.8 and TM-20.9.

REPORT FOR MATHEMATICS --- KAILUA HIGH SCHOOL

TREND CRITERIA= 5 PERIOD 11 20 JAN 1969

UNIT	TREND BY PERIOD											PERCENT CORRECT BY PERIOD										
	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	9	10	11
1	.	+	+	.	.	.	+	.	.	+	.	8	59	68	65	68	63	76	77	73	79	81
2	.	+	+	+	.	−	+	+	.	.	.	10	28	47	58	61	55	63	69	71	71	75
3	.	+	.	+	+	.	.	+	.	−	+	4	18	16	38	63	59	59	73	74	65	72
4	.	.	.	.	+	+	+	.	.	.	+	6	6	6	7	16	36	52	55	57	54	68
5	.	.	.	.	.	.	+	+	+	.	+	2	1	2	3	4	6	13	27	46	45	59
6	.	.	.	.	.	.	.	.	.	+	.	3	3	5	4	7	5	9	12	15	48	47
7	.	.	.	.	.	.	.	.	.	+	+	6	4	6	7	6	8	11	15	13	33	47

Figure TM-20.5. Sample computer output of group achievement on instructional units for administration eleven and all previous administrations. Course is 11th grade trigonometry; weekly tests were more frequent than completion of units. Pretesting indicated little prior knowledge of subject.

REPORT FOR MATHEMATICS --- KAILUA HIGH SCHOOL

TREND CRITERIA= 1 PERIOD 10 20 JAN 1969

NAME	FORM	SCORE BY PERIOD 1	2	3	4	5	6	7	8	9	10
AOKI MICHAEL	34	1	4	2	2	2	3	4	6	4	4
BANKS KIRK	36	1	3	3	3	2	1	5	4	5	10
BARRA ROBERTA	30	0	4	1	2	7	7	7	7	9	8
BEST LYNNDA	23	0	4	4	5	4	7	5	8	6	6
BOHOWITZ KATHLEEN	37	4	1	2	4	2	2	5		5	10
BRANHAM JOHN	28	5	4	2	3	5	5	5	0	7	9
BREHM GAIL	24	0	2	1		3	2	3	4	4	6
BROWN CAROLE	35	0	3	7	6	4	3	2	11	6	13
BUDD NANCY	28	0	2	4	5	4	4	7	8	10	9
BUNDY BONNIE	27	4	4	1	4	5	6	6	5	11	9
BURKE GAIRETT	31	3	1	3	2	4	2	6	6		0
BURNETT DAVID	22	3	4	4	4	4	5	8	10	9	10
BURNETT PEGGY	29	0	3	3	6	6	5	9	6	7	10
BUTCHART DAVID	24		0	0	6	6	6	7	11	9	11
BUTLER STEVEN	26	6	4	3	2	6	8	7	9	8	12
CHANG LABAN	24	5	0	2	1	5	2	4	1	4	11
CHINA MYRON	27		3	5		3	4	4	8	5	8
COSTA LINDA	21	0	0	2	2	4	3	5	6	4	8
CUSTER JACK	32	1	1	0	3	1	3	5	0	2	2
CUSTER KITTEE	29		2	3	3	3	4	3	5	5	7
DECOSTA LAVERNE	24X	3	1	1	3		3	2	1		7
DOWD STEVEN	25	5	2	3	4	5	1	5	6	8	8
DOZIER DOUG	32	0	2	3	5	6	5	7	7	8	7
FONG JEANETTE	33	1	3	5	4	6	8	7	8	8	12
GAUEN NANCY	35	1	4	5	3	3	2	3		7	7
GOYA BRUCE	29	0	2	3	3	8	6	3	7	7	6
GRAHAM MARC ANTONY	30	2	4	2	2	2	6	2	5	7	4
HAMILL MARION	37	0	2	3	4		4	6	6	5	10
JANSEN DEBRA	29	0	0	2	2	6	3	5	2	2	5
HARWELL CAROL	21	1	2	2	2	2	5	6	7	7	8
HEEKIN ROBERT	33	2	3	3	1	1	7	4	5		7
HOLMES DAVID	31	0	0	1			0	3	2	4	6

Figure TM-20.6. Sample computer output at each test adminitration: students' total scores for administration 10 and all previous administrations.

```
ELAM GARY                                KA442
PERIOD  3 - 12 DEC 1968
TOTAL SCORE IS  9                        PAST
                                         DATA
  PC A       PC A       PC A            PD  SC
  101 +                                  1   4
  102 +                                  2   2
  202 +                                  3   5
  204 +                                  4   4
  401 +                                  5   5
  403 +                                  6   8
  507 +                                  7   6
  901 -                                  8   9
 1001 -                                  9
 1004 -                                 10
 1103 -                                 11
 1105 +                                 12
 3042 +                                 13
 5042 -                                 14
                                        15
                                        16
                                        17
                                        18
                                        19
                                        20
```

```
HARWELL CAROL                            KA442
PERIOD  8 - 12 DEC 1968
TOTAL SCORE IS  7                        PAST
                                         DATA
  PC A       PC A       PC A            PD  SC
  102 +                                  1   1
  104 +                                  2   2
  404 +                                  3   2
  503 +                                  4   2
  507 +                                  5   2
  902 -                                  6   5
  905 -                                  7   6
 1003 -                                  8   7
 1103 -                                  9
 1106 -                                 10
 2052 +                                 11
 3011 -                                 12
 3041 +                                 13
 6012 -                                 14
                                        15
                                        16
                                        17
                                        18
                                        19
                                        20
```

Figure TM-20.7. Samples of individual student reports distributed after each test administration. PC is performance criterion number (objective number; a student may refer to his list of objectives). A is answer: + (right); - (wrong). PD is period (test administration) and SC is score.

ACHIEVEMENT PROFILES FOR HS420 --- AVERAGES BY MONITOR PERIOD

SG 4. STUDENTS IN HS420 WHO ANSWERED 11 TO 15 QUESTIONS ON PRETEST (N=38)
AND
QG 6. PERFORMANCE OBJECTIVES FROM UNIT FIVE INCLUDED IN PROFILES

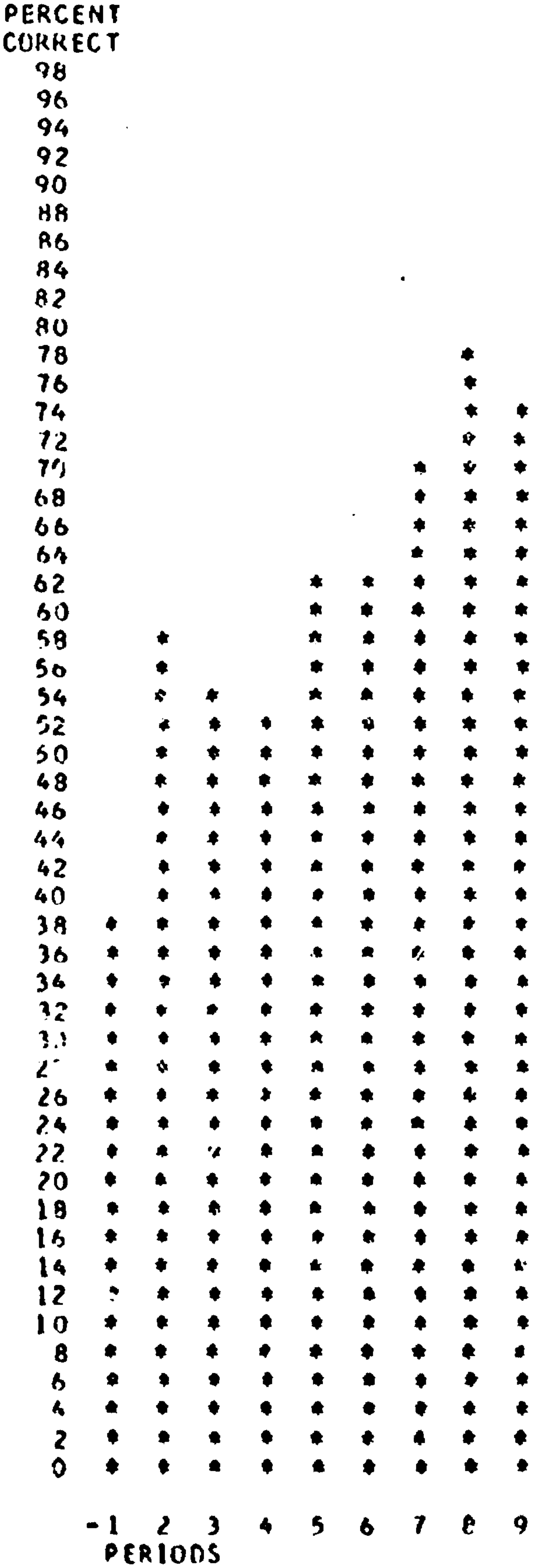


Figure TH-20.8. Sample computer output available when desired: profile for group (or subgroup) achievement on one objective or one instructional unit.

TABLE TM-20.8

Item Analysis: Criterion Scores by Time Intervals (days measured from completion of instruction). Appropriate items collapsed across intervals.

Student	Interval -150 / -15 (Pre-)	 0 / 60 (Post)	 61 / 150 (Retention)
18	03	38	59
63	01	60	60
79	14	82	84
109	07	27	42

TABLE TM-20.9

Item Analysis: Difficulty Indices by Time Intervals (days measured from completion of instruction). Appropriate students collapsed across intervals.

Item	Interval -150 / -15 (Pre)	 0 / 60 (Post)	 61 / 150 (Retention)
		Difficulty	
33	.07	.65	.94
51	.12	.67	.56
119	.09	.32	.69
174	.09	.29	.16
		Discrimination	
33	-.42	.48	-.14
51	-.15	.43	.72
119	-.12	.83	.48
174	.06	.14	-.13

VII. Implementation

Three Alternatives

CAM would prove valuable as one major form of evaluation of urban education projects. However, there could be considerable variation in the extent to which the state department (or other agency) provides expertise and assistance in evaluation. The degree of involvement of an agency external to a project is, in general, positively related to the quality and generalizability of the data collected. It would specifically affect comparability measure of effectiveness between projects concerned with the same objectives. Three alternatives are suggested below (see Table TM-20.9) for allocating authority for the various steps in the CAM procedure.

Handbook. Central to any plan to implement CAM would be the preparation of a handbook. This would provide a detailed description of the CAM procedure and its specific application to urban education projects. An example of this type of handbook is the Assessment and Evalutation--Title I--ESEA, which was developed by the New York State Education Department.

I. Project autonomy. With the aid of the handbook, a project director could design a CAM monitoring procedure for his project, including objectives, actual test items, and monitoring schedule. This preparation could be required as part of the proposal, or before funds were released, and would be critiqued at the State level. Individual projects would be responsible for the collection and analysis of data, and the reporting of results.

TABLE TM-20.10

Alternatives for State Level Control of Components of CAM Evalution

Component	Alternative I	Alternative II	Alternative III
Cam handbook	X	X	X
Objectives		X	X
Instruments		X	X
Design			X
Collecting data			X
Analysis of data			X
Reporting			X

One benefit of such intimate involvement with the evaluation is the increased likelihood that project directors would correct weaknesses as they became apparent. Decisions about pacing and sequencing would rest on the working knowledge gained from defining objectives and evaluation instrument, and from following a systematic procedure during the year.

A potentially serious limitation of this alternative is the variation across projects in the quality and uniformity of instruments, and in the quality and uniformity of collection and anlysis of the data.

<u>II</u>: <u>State specified objectives and items</u>. A State level agency would be responsible not only for the handbook, but for clearly articulating the objectives and associated items for a project or series of related projects. This would probably involve the use of a computerized item bank, to facilitate the distribution of the items in the form of a set of CAM monitors. Since this process takes skilled staff and time, materials would be developed gradually; for the most pay-off, projects affecting the largest number of students would be chosen first.

Possible advantages of alternative II over I include the probability that objectives will be better clarified and items more valid. Also, uniformity of items from project to project for evaluating similar objectives allows clearer comparisons of project success in achieving objectives.

Part of the validity and comparability of information is a function of the reliability and uniformity of the design, data collection, analysis, and reporting. Since these elements are left in the hands of the individual project's personnel, a <u>relative</u> comparison of projects is less valid. It must be noted, however, that validity of data collection is very much a function of the training and carefulness of the person actually administering the monitors, given a good evaluation design.

III. State control over all components. Alternative III calls for a comprehensive evaluation to be made of each project, with administrative responsibility separate from the project. The evaluation agency would prepare and operate all major components of the evaluation, and would continuously report results to both the project and the State Education Department. Specifically, the agency would write a CA' handbook, assist the project in articulating its objectives, develop items, specify the design, collect and analyze data, and report results.

The resources needed for the evaluation agency would be developed gradually over a one-and-a-half to a five-year period, depending on the level of financial support. The development would begin with a sample of projects with high priorities e.g. reading for non-English speaking students. Many of the resources developed for the reading evaluation (e.g. staff, computer programs, data collection techniques) would be directly usable with othr projects. Only additional objectives, items, and staff would then need to be developed.

Budget and Timetable

Table TM-20.11 presents some approximate costs for the three alternatives presented above. Since Alternative III encompasses the first two alternatives, the table presents the costs cumulatively. The estimates are itemized, to make it possible to adjust them to more probable rates within the Education Department. It should be noted that these costs are based on ten monitor administrations in fifteen projects of the same type.

For a one-year project, it might be feasible to think in terms of five monitor administrations: one each quarter and one at the beginning of the year. However, it would be much more difficult to correct the plan of a

TABLE TH-20.11

Cost Estimates for Fifteen Projects of Same Type

Alternative	Estimate Procedure			
	Category	Itemization		Cost
I	Handbook			$ 15 000.[a]
		2 writers, 2 months	$ 5 000.	
		secretary, 3 months	2 000.	
		printing, 10 000 cps.	5 000.	
		Overhead		
II	Objectives and Items			40 000.[a,b]
		2 wk. wkshp. exons.	5 000.[b]	
		2 people/proj. @ $400.	12 000.[b]	
		1 administrator, 3 mos.	5 000.[b]	
		2 curric. consult., 2 mos.	2 500.[b]	
		3 curric. consult., 2 wks.	1 500.[b]	
		2 secretaries, 3 mos.	3 000.[b]	
		overhead, duplicating, telephone, etc.		
	Computer based test forms			10 000.
		keypunching, 2 items/5 min. @ $4./hr.; 160 hrs. for 4000 items[c]	800.[a,b]	
		computer time (development, print monitors on mimeo)	2 000.[b]	
		sampling programmer	1 000.[a]	
		item-bank programmer	2 500.[a]	
		running technician	1 000.[b]	
	Evaluation design	$200./project		3 000.[b]
	(Each additional project of same type)			(1 000.)
	(Projects of another type; per project)			(5 000.)[d]

(Table TM-20.11, Continued)

III	Revise and further develop computer programs for data bank, analysis, and reporting		25 000.[a]
	computer time; 1/2 hrs./da.; 5 da./wk.; 12 wks. (30 hrs. @ $300.)	10 000.	
	two programmers for 12 wks.	8 000.	
	Data collection and anlysis		40 000.[b]
	15 projects; 1 test/4 wks. = 1 test/proj./day		
	1 data collector - full time	12 000.	
	Training (1 mo.) - Trainer Other expenses	1 500.	
	Punch data (e.g. Digitek) 500 Ss/proj./10 tests @ 5¢	4 000.	
	Analysis & reportingi 15 min./proj./testing @ $300./hr.	10 500.	
	Reports to projects	300.	
	mailing $1./testing		
	handling $1./testing		
	State analysis of data 1/2 da./test/proj.=75 days @ $50.	3 750.	

Summary

I	Handbook		$ 15 000.
II	Project autonomy		70 000.
	Handbook	$ 15 000.	
	Objectives and items	40 000.	
	Test forms	10 000.	
	Evaluation design	3 000.	
III	State control over all components		135 000.
	Alternative II	$ 70 000.	
	Computer programs	25 000.	
	Collection and Analysis	40 000.	

Note: Costs for alternatives are cumulative (see Summary). Totals are round figures, not exact summations.

a one time costs

b repeating costs

c 1000 items in common, 200 unique items for each project.

d depends somewhat on number of projects in the type

course if half a semester were allowed to go by without feedback. The chief advantage of the use of CAM for state level evaluation is the data that are fed back to the projects for their own adjustment of strategy before the course is over. The cost estimates reflect this philosophy.

Table TM-20.12 presents the tightest practical schedule for executing each component of CAM. The following points must be held in mind while interpreting the timetable.

1) The handbook is scheduled here to be completed before the construction of objectives and items. However, this is not necessary; state department personnel have many resources for constructing a curriculum in behavioral terms, and writing valid items for measuring specific objectives. The handbook should be completed in time to aid the writers of incoming proposals. It is obviously more important for those who may construct curricula and tests without the direct supervision of the state department (Alternative I).

2) Time allotted for each task minimum, and assumes fulltime, efficient attention; each task is scheduled at the latest possible time it could be done. Needless to say, it is adviseable to leave extra time for unforeseen delays. Time allowances must also be adjusted if the personnel involved are working on other assignments simultaneously. Some tasks may be completed sooner if more personnel are available.

3) The timetable is related only to the tasks which need to be completed before actually starting to use CAM; it is independent of who does the work; i.e., it is not directly related to the Alternatives suggested above. However, it is inevitable that coordinating the efforts of more people on more levels will take more time.

TABLE TM-20. 12

Timetable for Completion of Components of CAM

TASK	Lead time needed in months						Start of Course
	6	5	4	3	2	1	
Handbook							
writing							
printing							
distribution							
Objectives & items							
intensive development (workshop)							
editing							
keypunching							
processing in item bank (monitor forms on mimeos)							
mailing mimeos							
project running of mimeos and collating							
Item bank revisions & sampling procedures							
Evaluation design (1-2 days per project)							
Computer program revisions (Data bank, Analysis, Report)							
Data collectors' training							

Practical Considerations

Planning CAM early. To be most effective, CAM must be planned when a course is planned, in fact, CAM is inseparable from the content and structure of a course. Time for preparation, funds, and staff must be integral parts of the evaluation planned in a proposal. It would be advisable to require such a plan as part of every proposal, whether or not it was to be modified later with the help of State consultants.

Cost of CAM. If a small expenditure on evaluation doesn't yield reliable data, it is very expensive evaluation. If five to ten percent of project funds were allotted to well-designed evaluation, it could allow total dollars spent in education to have substantially more impact. There are two major observations to make about the cost of CAM. 1) If the State were to coordinate personnel from similar projects for objective and item writing, the additional cost for CAM would represent a fraction of the funds already allocated to traditional evaluation. 2) The basic preparation for CAM is the writing of objectives in clear, testable, behavioral terms, and constructing items to measure specific objectives. Obviously, just preparing to use CAM will vastly improve the quality of a project, quite apart from actual data collection or reporting results. It is questionable whether this expense should be charged to "evaluation," or to "curriculum construction." If the curriculum has been well specified, then setting up a CAM evaluation procedure is relatively easy. When the benefits of feedback to both students and teachers are considered, the "cost of evaluation" is small indeed.

Writing objectives and items. The job of specifying a curriculum for a course or project in behavioral terms, and writing questions to measure

those objectives, is substantial. One strategy would be for a team of teachers and curriculum specialists to work together to define the objectives of the project. Staff from similar projects could develop a bank of objectives and items, and a set of tests could then be tailored to a specific project by stratified sampling. A feasible way to expedite this whole process is to hold a workshop for several weeks during the summer, with curriculum people and/or directors of similar projects grouped together. This plan could be considered a desirable modification to Alternative I and has been included in the budget for Alternative II.

Collection of all possible raw data. The ideal use of CAM would be to monitor all students at each administration, and to do this often, at least every two weeks. This would provide the maximum amount of data: it would be directly useful to the project (CAM imposes more than average structure on data collection; therefote the data are more valuable); teachers would find it a satisfactory substitute for their own test programs, and it could be appropriately summarized for state-level collation with data from similar projects.

Choosing an alternative. The three plans for implementation of CAM each have particular merit. However, from the point of view of pure evaluation at the state level, Alternative III has the most validity built in. Ultimately, the cost for state evaluation of state-funded projects is a matter of which pocket the money comes out of, rather than from whom.

No one alternative need be selected on a statewide basis, however. The choice could be a matter of excellence of design for evaluation in a proposal, as critiqued by the state; or project directors might express a preference; or certain types of projects may be better suited to one

plan than another. The most important consideration is that the evaluation procedure should serve the project locally as feedback on which to base "real time" decisions, and also serve the state-level needs in project evaluation.

GENERAL BIBLIOGRAPHY

Bond, G. L. and Dykstra, R. The Cooperative research program in first-grade reading instruction. Reading Research Quarterly, 2, 4, 1-140, 1967.

Bloom, B. S. Taxonomy of education objectives: cognitive domain. New York: David McKay, 1956.

Cronbach, L. J. Evaluation for course improvement. Teachers College Record, 1963, 64, 672-683.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. Gagne (Ed.), Learning and individual differences. Columbus, Ohio: Merrill Books, 1967.

Harman, H. H., Helm, C. E., & Loye, D. E. Computer assisted testing. Princeton: Educational Testing Service, 1968.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Menlo Park, Ca.: Addison-Wesley, 1968.

Mager, R. F. Preparing instructional objectives. Palo Alto, Ca.: Feron Publishers, 1962.

New York State, Assessment and evaluation handbook-Title I-ESEA. Albany, New York: The University of the State of New York, The State of New York, The State Education Department, Office of Research and Evaluation, 1968.

New York State, Guidelines for New York State urban education program, Albany, New York: State Education Department, Office of Urban Education, 1968.

Stake, R. E. (Ed.) Perspectives of curriculum evaluation. Chicago: Rand McNally, 1967.

CAM BIBLIOGRAPHY

Journal Articles

Gorth, W. and Grayson, A. A program to compose and print tests for instructional testing using item sampling. I. Educational and psychological measurement, Spring, 1969.

Gorth, W., Grayson, A., Popejoy, L. and Stroud, T. A tape-based data bank from educational research or instructional testing using longitudinal item sampling. II. Educational and psychologcal measurement, Spring, 1969.

Gorth, W., Grayson, A. and Stroud, T. A computer program to tabulate and plot achievement profiles of longitudinal achievement testing using item sampling. III. Educational and psychological measurement, Spring, 1969.

Gorth, W., Grayson, A. and Lindeman, R. A computer program to evaluate item performance by internal and external criteria in a longitudinal testing program using item sampling. IV. Educational and psychological measurement, Spring, 1969.

Gorth, W., Allen, D., Popejoy, L. and Stroud, T. A comparison of comprehensive versus unit pretesting and students' achievement, Psychology in the Schools (In Press).

Lindeman, R., Gorth, W. and Allen, D. Item analysis in an item sampling case. Journal of Educational Measurement (In Press).

Professional Meetings

Gorth, W. (Organizer) Comprehensive Random Achievement Monitoring. Symposium presented at the meeting of the American Educational Research Association, Los Angeles, February, 1969.

Gorth, W. and Allen, D. A new design for evaluation in mathematics education. Paper to be presented at the meeting of the National Council of Teachers of Mathematics, Minneapolis, April, 1969.

Lindeman, R., Gorth, W. and Allen, D. Item analysis in an item sampling case. Paper presente

Technical Memoranda

Number	Title	Author
AR-1	First Annual Report	Gorth
TM-2	Description of courses monitored by Project CRAM	Gorth & Popejoy
TM-3	Monitoring schedules developed for research by Project CRAM	Gorth, Stroud, & Knight
TM-4	The relation of repeated, comprehensive pretesting and students' achievement	Gorth, Allen, Popejoy, & Stroud
TM-5	A comparison of comprehensive versus unit pretesting and students' achievement	Popejoy, & Stroud
TM-6	The evaluation of item performance in an item sampling case	Lindeman, Gorth, & Allen
TM-7	Computer-Based, instructional-testing data bank	Popejoy, Gorth, Grayson & Stroud
TM-8	Separate analyses of regression	Stroud & Gorth
TM-9	Education innovations monitored by Project CRAM	Gorth
TM-10	Longitudinal comprehensive achievement monitoring in science education	Gorth & Allen
TM-11	A computer program to evaluate item performance by internal and external criteria in a longitudinal testing program using item sampling	Gorth, Grayson & Lindeman
TM-12	A computer program to tabulate performance profiles of logitudinal performance testing using item sampling	Gorth, Grayson, & Stroud
TM-13	The Project CRAM data bank for 1967-1968	Gorth
TM-14	A computer program to compose and print tests for instructional testing using item sampling	Gorth
TM-15	Investigating a linear model of learning in ninth grade algebra	Stroud & Gorth
TM-16	Analysis of the Project CRAM data for 1967-1968	Gorth & Pinsky
TM-17	Monitoring schedules developed for research; 1968-1969	Pinsky & Gorth
TM-18	Demographic, aptitude, & attitude surveys of the students, teachers, and shools in Project CRAM	Gorth & Pinsky
AR-2	Second Annual Report to the Charles F. Kettering Foundation	Allen & Gorth

APPENDIX

Project CAM Staff

Principal Investigator:	Dwight W. Allen
Director:	William P. Gorth
Research Associates:	David Evans Donald DeLay Kenneth Knight Richard Lindeman Lawrence Wightman
Research Assistants:	Paul Pinsky Lee Popejoy Nathaniel Sims George Worle
Statistical Consultants:	Thomas Stroud Richard Snow
Computer Programmer:	Aram Grayson